

**WHAT IS CLAIMED IS:**

1. A method of processing text data, comprising the steps of:  
inputting text data;  
5 parsing the text data into word candidates;  
removing predetermined words from the word candidates;  
specifying an area of a predetermined text database; and  
determining a specific area occurrence value of each of the word candidates in the  
specified area in the predetermined text database.  
10
2. The method of processing text data according to claim 1 wherein the specified area is a  
header area.
3. The method of processing text data according to claim 2 wherein the specific area  
15 occurrence value is determined according to a following equation:  
  
the specific area occurrence value =  
a number of documents including the word candidate in the header area /  
a number of documents including the word candidate in an entire portion of the  
20 predetermined text database.
4. The method of processing text data according to claim 1 wherein the specified area is a  
summary area.
- 25 5. The method of processing text data according to claim 4 wherein the specific area  
occurrence value is determined according to a following equation:  
  
the specific area occurrence value =  
a number of documents including the word candidate in the summary area /  
30 a number of documents including the word candidate in an entire portion of the  
predetermined text database.

6. The method of processing text data according to claim 1 wherein the specified area is a combination of a header area and a summary area.

7. The method of processing text data according to claim 6 wherein the specific area  
5 occurrence value is determined according to a following equation:

the specific area occurrence value =  
a number of documents including the word candidate in either one of the  
summary area and the header area /  
10 a number of documents including the word candidate in an entire portion of the  
predetermined text database.

8. The method of processing text data according to claim 6 wherein the specific area  
occurrence value is determined according to a following equation:

15 the specific area occurrence value =  
(a number of documents including the word candidate in the header area /  
a number of documents including the word candidate in an entire portion of the  
predetermined text database) +  
20 (a number of documents including the word candidate in the summary area /  
a number of documents including the word candidate in an entire portion of the  
predetermined text database)

9. The method of processing text data according to claim 1 further comprising an  
25 additional step of determining a search word significance value based upon a following  
equation:

the search word significance value =  
a corresponding predetermined word weight X  
30 the specific area occurrence value,  
wherein the corresponding predetermined word weight is  $\log$  (a total number of  
documents/ the number of documents in which the word candidate occurs).

10. The method of processing text data according to claim 1 further comprising an additional step of:

5           determining a search word significance value based upon a following equation:

                  the search word significance value =

                          a corresponding predetermined word weight X

                          the specific area occurrence value X

10                   a number of occurrences of the word candidate within the text data.

11. The method of processing text data according to claim 1 further comprising additional steps of:

15           selecting search words from the word candidates based upon the specific area occurrence value; and

                  extracting sentences from the predetermined text database based upon the selected search words.

20   12. The method of processing text data according to claim 1 further comprising an additional step of selecting keywords from the word candidates based upon the specific area occurrence value.

13. The method of processing text data according to claim 1 further comprising additional steps of:

25           selecting keywords from the word candidates based upon the specific area occurrence value; and

                  generating a summary from the predetermined text database based upon the selected keywords.

30

14. The method of processing text data according to claim 1 further comprising additional steps of:

selecting classification keywords from the word candidates based upon the specific area occurrence value; and

classifying the predetermined text database based upon the selected classification keywords.

5

15. The method of processing text data according to claim 1 further comprising additional steps of:

determining a first text database occurrence value of the word candidates in a first text database;

10 determining a second text database occurrence value of the word candidates in a second text database;

determining a database occurrence value based upon the first text database occurrence value and the second text database occurrence value in a predetermined manner;

15 selecting search words from the word candidates based upon in part the database occurrence value; and

extracting sentences from a predetermined text database based upon the selected search words.

20 16. The method of processing text data according to claim 15 wherein the database occurrence value is determined by a following equation:

the database occurrence value =

25 
$$\frac{(\text{the second text database occurrence value} / \text{a total number of documents in the second text database}) - (\text{the first text database occurrence value} / \text{a total number of documents in the first text database})}{1}$$

17. The method of processing text data according to claim 15 wherein the database occurrence value is determined by a following equation:

30

the database occurrence value =

(the second text database occurrence value /  
a total number of documents in the second text database) /  
(the first text database occurrence value /  
a total number of documents in the first text database).

5

18. The method of processing text data according to claim 15 further comprising an additional step of determining a search word significance value based upon a following equation:

10 the search word significance value =

the corresponding predetermined word weight X  
the database occurrence value,

wherein the corresponding predetermined word weight is  $\log$  (a total number of documents/ the number of documents in which the word candidate occurs).

15

19. A method of processing text data, comprising the steps of:

inputting text data;

parsing the text data into word candidates;

removing predetermined words from the word candidates;

20 determining a first text database occurrence value of the word candidates in a first text database;

determining a second text database occurrence value of the word candidates in a second text database;

25 determining a database occurrence value based upon the first text database occurrence value and the second text database occurrence value in a predetermined manner;

selecting search words from the word candidates based upon in part the database occurrence value; and

30 extracting sentences from a predetermined text database based upon the selected search words.

20. The method of processing text data according to claim 19 wherein the database occurrence value is determined by a following equation:

the database occurrence value =

$$\begin{aligned} & \text{(the second text database occurrence value /} \\ & \text{a total number of documents in the second text database) -} \\ & \text{(the first text database occurrence value /} \\ & \text{a total number of documents in the first text database).} \end{aligned}$$

21. The method of processing text data according to claim 19 wherein the database occurrence value is determined by a following equation:

the database occurrence value =

$$\begin{aligned} & \text{(the second text database occurrence value /} \\ & \text{a total number of documents in the second text database) /} \\ & \text{(the first text database occurrence value /} \\ & \text{a total number of documents in the first text database).} \end{aligned}$$

22. The method of processing text data according to claim 19 further comprising an additional step of determining a search word significance value based upon a following equation:

the search word significance value =

$$\begin{aligned} & \text{the corresponding predetermined word weight X} \\ & \text{the database occurrence value,} \end{aligned}$$

wherein the corresponding predetermined word weight is  $\log(\text{a total number of documents} / \text{the number of documents in which the word candidate occurs})$ .

23. A computer program for processing text data, performing the tasks of:

- inputting text data;
- parsing the text data into word candidates;
- removing predetermined words from the word candidates;
- specifying an area of a predetermined text database; and

determining a specific area occurrence value of each of the word candidates in the specified area in the predetermined text database in a predetermined manner.

24. The computer program for processing text data according to claim 23 wherein the  
5 specified area is a header area.

25. The computer program for processing text data according to claim 24 wherein the specific area occurrence value is determined according to a following equation:

10 the specific area occurrence value =  
a number of documents including the word candidate in the header area /  
a number of documents including the word candidate in an entire portion of the  
predetermined text database.

15 26. The computer program for processing text data according to claim 23 wherein the specified area is a summary area.

27. The computer program for processing text data according to claim 26 wherein the specific area occurrence value is determined according to a following equation:

20 the specific area occurrence value =  
a number of documents including the word candidate in the summary area /  
a number of documents including the word candidate in an entire portion of the  
predetermined text database.

25 28. The computer program for processing text data according to claim 23 wherein the specified area is a combination of a header area and a summary area.

29 The computer program for processing text data according to claim 28 wherein the  
30 specific area occurrence value is determined according to a following equation:

the specific area occurrence value =

a number of documents including the word candidate in either one of the  
summary area and the header area /

a number of documents including the word candidate in an entire portion of the  
predetermined text database.

5

30. The computer program for processing text data according to claim 28 wherein the  
specific area occurrence value is determined according to a following equation:

the specific area occurrence value =

10 (a number of documents including the word candidate in the header area /  
a number of documents including the word candidate in an entire portion of the  
predetermined text database) +  
(a number of documents including the word candidate in the summary area /  
a number of documents including the word candidate in an entire portion of the  
15 predetermined text database).

31. The computer program for processing text data according to claim 23 further  
comprising an additional task of determining a search word significance value based upon  
a following equation:

20

the search word significance value =

a corresponding predetermined word weight X  
the specific area occurrence value,

wherein the corresponding predetermined word weight is  $\log$  (a total number of  
25 documents/ the number of documents in which the word candidate occurs).

32. The computer program for processing text data according to claim 23 further  
performing an additional task of determining a search word significance value based upon  
a following equation:

30

the search word significance value =

a corresponding predetermined word weight X



the specific area occurrence value  $X$   
a number of occurrences of the word candidate within the text data.

33. The computer program for processing text data according to claim 23 further  
5 performing additional tasks of:  
    selecting search words from the word candidates based upon the specific area  
occurrence value; and  
    extracting sentences from the predetermined text database based upon the selected  
search words.
- 10 34. The computer program for processing text data according to claim 23 further  
performing an additional task of selecting keywords from the word candidates based upon  
the specific area occurrence value.
- 15 35. The computer program for processing text data according to claim 23 further  
performing additional tasks of:  
    selecting keywords from the word candidates based upon the specific area  
occurrence value; and  
    generating a summary from the predetermined text database based upon the  
20 selected keywords.
36. The computer program for processing text data according to claim 23 further  
performing additional tasks of:  
    selecting classification keywords from the word candidates based upon the  
25 specific area occurrence value; and  
    classifying the predetermined text database based upon the selected classification  
keywords.
37. The computer program for processing text data according to claim 23 further  
30 performing additional task of:  
    determining a first text database occurrence value of the word candidates in a first  
text database;

determining a second text database occurrence value of the word candidates in a second text database;

determining a database occurrence value based upon the first text database occurrence value and the second text database occurrence value in a predetermined  
5 manner;

selecting search words from the word candidates based upon in part the database occurrence value; and

extracting sentences from the predetermined text database based upon the selected search words.

10

38. The computer program for processing text data according to claim 37 wherein the database occurrence value is determined by a following equation:

the database occurrence value =

(the second text database occurrence value / a total number of documents in the second text  
15 database) –

(the first text database occurrence value / a total number of documents in the first text database).

39. The computer program for processing text data according to claim 37 wherein the  
20 database occurrence value is determined by a following equation:

the database occurrence value =

(the second text database occurrence value /  
a total number of documents in the second text database) /  
25 (the first text database occurrence value /  
a total number of documents in the first text database).

40. The computer program for processing text data according to claim 37 further  
performing an additional task of determining a search word significance value based upon  
30 a following equation:

the search word significance value =

the corresponding predetermined word weight X  
the database occurrence value,  
wherein the corresponding predetermined word weight is  $\log$  (a total number of  
documents/ the number of documents in which the word candidate occurs).

5

41. A computer program for processing text data, performing the tasks of:

inputting text data;

parsing the text data into word candidates;

removing predetermined words from the word candidates;

10 determining a first text database occurrence value of the word candidates in a first  
text database;

determining a second text database occurrence value of the word candidates in a  
second text database;

determining a database occurrence value based upon the first text database  
15 occurrence value and the second text database occurrence value in a predetermined  
manner;

selecting search words from the word candidates based upon in part the database  
occurrence value; and

20 extracting sentences from the predetermined text database based upon the selected  
search words.

42. The computer program for processing text data according to claim 41 wherein the  
database occurrence value is determined by a following equation:

25 the database occurrence value =

(the second text database occurrence value /  
a total number of documents in the second text database) –  
(the first text database occurrence value /  
a total number of documents in the first text database).

30

43. The computer program for processing text data according to claim 41 wherein the  
database occurrence value is determined by a following equation:

the database occurrence value =

$$\frac{\begin{array}{l} \text{(the second text database occurrence value /} \\ \text{a total number of documents in the second text database) /} \\ \text{(the first text database occurrence value /} \\ \text{a total number of documents in the first text database).} \end{array}}{5}$$

44. The computer program for processing text data according to claim 41 further comprising an additional step of determining a search word significance value based upon  
10 a following equation:

the search word significance value =

$$\frac{\text{the corresponding predetermined word weight X}}{\text{the database occurrence value,}}$$

15 wherein the corresponding predetermined word weight is  $\log$  (a total number of documents/ a number of documents including the word candidate in an entire portion of the predetermined text database).

45. A apparatus for processing text data, comprising:

20 an input unit for inputting text data;  
a search word selection unit connected to said input unit for parsing the text data into word candidates, said search word selection unit removing predetermined words from the word candidates;

an area specification unit for specifying an area of a predetermined text database;

25 and

a specific area occurrence determination unit connected to said search word selection unit and said area specification unit for determining a specific area occurrence value of each of the word candidates in the specified area in the predetermined text database.

30

46. The apparatus for processing text data according to claim 45 wherein the specified area is a header area.

47. The apparatus for processing text data according to claim 46 wherein said specific area occurrence determination unit determines the specific area occurrence value according to a following equation:

5

the specific area occurrence value =

a number of documents including the word candidate in the header area/

a number of documents including of the word candidate in an entire portion of the predetermined text database.

10

48. The apparatus for processing text data according to claim 45 wherein the specified area is a summary area.

49. The apparatus for processing text data according to claim 48 wherein said specific area occurrence determination unit determines the specific area occurrence value according to a following equation:

15

the specific area occurrence value =

a number of documents including the word candidate in the summary area /

a number of documents including the word candidate in an entire portion of the

20

predetermined text database.

50. The apparatus for processing text data according to claim 45 wherein the specified area is a combination of a header area and a summary area.

51. The apparatus for processing text data according to claim 50 wherein said specific area occurrence determination unit determines the specific area occurrence value according to a following equation:

25

the specific area occurrence value =

a number of documents including the word candidate in either one of the

30

summary area and the header area /

a number of documents including the word candidate in an entire portion of the predetermined text database.

52. The apparatus for processing text data according to claim 50 wherein said specific area occurrence determination unit determines the specific area occurrence value according to a following equation:

5 the specific area occurrence value =

$$\frac{\text{(a number of documents including the word candidate in the header area / a number of documents including the word candidate in an entire portion of the predetermined text database)} + \text{(a number of documents including the word candidate in the summary area / a number of documents including the word candidate in an entire portion of the predetermined text database)}}{2}$$

10

53. The apparatus for processing text data according to claim 45 wherein said search word selection unit further determines a search word significance value based upon a following equation:

15 the search word significance value =

$$\frac{\text{a corresponding predetermined word weight X the specific area occurrence value,}}{\text{wherein the corresponding predetermined word weight is log (a total number of documents/ the number of documents in which the word candidate occurs.)}}$$

20

54. The apparatus for processing text data according to claim 45 wherein said search word selection unit further determines a search word significance value based upon a following equation:

25 the search word significance value =

$$\frac{\text{a corresponding predetermined word weight X the specific area occurrence value X a number of occurrences of the word candidate within the text data.}}{2}$$

30 55. The apparatus for processing text data according to claim 45 further comprising a text selection unit connected to said specific area occurrence determination unit for selecting search words from the word candidates based upon the specific area occurrence value, said

text selection unit extracting sentences from the predetermined text database based upon the selected search words.

56. The apparatus for processing text data according to claim 45 further comprising a  
5 keyword extraction unit connected to said specific area occurrence determination unit for selecting keywords from the word candidates based upon the specific area occurrence value.

57. The apparatus for processing text data according to claim 45 further comprising:  
10 a keyword extraction unit connected to said specific area occurrence determination unit for selecting keywords from the word candidates based upon the specific area occurrence value; and  
a summary generation unit connected to said keyword extraction unit for generating a summary from the predetermined text database based upon the selected  
15 keywords.

58. The apparatus for processing text data according to claim 45 further comprising:  
a classification keyword selection unit connected to said specific area occurrence  
determination unit for selecting classification keywords from the word candidates based  
20 upon the specific area occurrence value; and  
a classification unit connected to said classification keyword selection unit for classifying the predetermined text database based upon the selected classification  
keywords.

25 59. The apparatus for processing text data according to claim 45 further comprising:  
a database occurrence determination unit connected to said search word selection unit for determining a first text database occurrence value of the word candidates in a first text database and a second text database occurrence value of the word candidates in a second text database, said database occurrence determination unit further determining a  
30 database occurrence value based upon the first text database occurrence value and the second text database occurrence value in a predetermined manner, wherein said search

word selection unit selects search words from the word candidates based upon in part the database occurrence value; and

a text selection unit connected to said search word selection unit for extracting sentences from the predetermined text database based upon the selected search words.

5

60. The apparatus for processing text data according to claim 59 wherein said database occurrence determination unit determines the database occurrence value based upon a following equation:

10 the database occurrence value =

$$\frac{\text{(the second text database occurrence value / a total number of documents in the second text database)} - \text{(the first text database occurrence value / a total number of documents in the first text database)}}{1}$$

15

61. The apparatus for processing text data according to claim 59 wherein said database occurrence determination unit determines the database occurrence value based upon a following equation:

20 the database occurrence value =

$$\frac{\text{(the second text database occurrence value / a total number of documents in the second text database)} / \text{(the first text database occurrence value / a total number of documents in the first text database)}}{1}$$

25

62. The apparatus for processing text data according to claim 45 wherein said search word selection unit further determines a search word significance value based upon a following equation:

30 the search word significance value =

$$\frac{\text{the corresponding predetermined word weight} \times \text{the database occurrence value}}{1}$$



wherein the corresponding predetermined word weight is  $\log$  (a total number of documents/ the number of documents in which the word candidate occurs).

63. A apparatus for processing text data, comprising:

- 5           an input unit for inputting text data;
- a search word selection unit connected to said input unit for parsing the text data into word candidates, said search word selection unit removing predetermined words from the word candidates;
- a database occurrence determination unit connected to said search word selection
- 10       unit for determining a first text database occurrence value of the word candidates in a first text database and a second text database occurrence value of the word candidates in a second text database, said database occurrence determination unit further determining a database occurrence value based upon the first text database occurrence value and the second text database occurrence value in a predetermined manner, wherein said search
- 15       word selection unit selects search words from the word candidates based upon in part the database occurrence value; and
- a text selection unit connected to said search word selection unit for extracting sentences from the predetermined text database based upon the selected search words.

20       64. The apparatus for processing text data according to claim 63 wherein said database occurrence determination unit determines the database occurrence value based upon a following equation:

the database occurrence value =

$$\begin{aligned}
 & \text{(the second text database occurrence value /} \\
 25 \quad & \text{a total number of documents in the second text database) -} \\
 & \text{(the first text database occurrence value /} \\
 & \text{a total number of documents in the first text database).}
 \end{aligned}$$

65. The apparatus for processing text data according to claim 63 wherein said database

30       occurrence determination unit determines the database occurrence value based upon a following equation:

the database occurrence value =

$$\frac{(\text{the second text database occurrence value} / \text{a total number of documents in the second text database})}{(\text{the first text database occurrence value} / \text{a total number of documents in the first text database})}.$$

5

66. The apparatus for processing text data according to claim 63 wherein said search word selection unit further determines a search word significance value based upon a following equation:

10 the search word significance value =

$$\frac{\text{the corresponding predetermined word weight}}{\text{the database occurrence value}},$$

wherein the corresponding predetermined word weight is  $\log(\text{a total number of documents} / \text{the number of documents in which the word candidate occurs})$ .

15